DOCUMENT RESUME

ED 318 788 TM 014 903

AUTHOR Anderson, Paul S.
TITLE Initial Experiences with Machine-Assisted Reconsiderative Test Scoring: A New Method for Partial Credit and Multiple Correct Responses.
PUB DATE Apr 90
NOTE 28p.; Paper presented at a joint session of the Annual Meetings of the American Educational Research Association (Boston, MA, April 16-20, 1990) and the National Council on Measurement in Education (Boston, MA, April 17-19, 1990).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS College Students; *Computer Assisted Testing; Educational Technology; Higher Education; High Schools; *Scoring; Secondary School Students; Test Construction; Test Items; *Test Scoring Machines
IDENTIFIERS Multi Digit Technique; *Multi Digit Tests; Partial Credit Model; *Reconsiderative Scoring

ABSTRACT

Initial experiences with computer-assisted reconsiderative scoring are described. Reconsiderative scoring occurs when student responses are received and reviewed by the teacher before points for correctness are assigned. Manually scored completion-style questions are reconsiderative. A new method of machine assistance produces an item analysis on a microcomputer that prints the actual word response or numeric answer. The teacher reviews the responses prior to allocating points via the keyboard. Computer-assisted reconsiderative scoring was first available in an experimental software package called RECON in the fall semester of 1989. Early experiences in university classes with about 250 students and secondary classes in one high school with the RECON package and a related software package--the MDT Educational Testing System--demonstrate a number of advantages, including: (1) the possibility of accepting multiple responses to one question; (2) enhanced numeric responses; (3) assessing multiple steps in responses; (4) allowing graphical responses; (5) improved feedback; and (6) links with computer managed learning and databases. Reconsiderative scoring could open a new dimension in educational measurement for teacher-generated and standardized assessments. Seven figures illustrate the discussion, and an appendix reports the development of the MDT program. A 38-item annotated bibliography is included. (SLD)

INITIAL EXPERIENCES WITH MACHINE-ASSISTED RECONSIDERATIVE TEST SCORING: A NEW METHOD FOR PARTIAL CREDIT AND MULTIPLE CORRECT RESPONSES

Paul S. Anderson
Illinois State University

Initial Experiences with Machine-Assisted Reconsiderative Test Scoring: A New Method for Partial Credit and Multiple Correct Responses

Paul S. Anderson, Ph.D.
Associate Professor, Department of Geography-Geology
Illinois State University, Normal, IL 61761
Vice President, Multi-Digit Technologies (MDT) Corp.
107 Broadway, Normal, IL 61761

ABSTRACT

Reconsiderative scoring occurs when student responses are received and reviewed by the teacher before points for correctness are assigned. Manually scored completion-style questions are reconsiderative. A new method of machine assistance produces on a microcomputer an item analysis that prints the actual word response or numeric answer. The teacher reviews the responses prior to allocating points via the keyboard. Initial experiences are reported. Pedagogical implications, including additional software capabilities to assess higher order learning, are presented. Reconsiderative scoring could open a new dimension in educational measurement for both teacher-generated and standardized assessments.

[Paper presented at the joint conference of the National Council for Measurement in Education (NCME) and the American Educational Research Association (AERA), Boston, MA, 16-19 April 1990.]

* * * * (A one-page synopsis is on page 26.)

I. Concepts and Definitions

A. Reconsiderative Scoring

Reconsiderative test scoring occurs when a student's response to a multiple choice or completion (fill-in-the-blank) question is reviewed by the teacher prior to the determination of correctness and point value. The reconsiderative method has been used for decades, even centuries, primarily in manual scoring of completion-style questions. Examples include cases where teachers award full or partial credit for misspelled responses, synonyms and numeric calculations with minor errors. Also, incomplete responses, such as not distinguishing between John Adams and John Quincy Adams or writing only "homo" instead of "homo sapiens," can require reconsiderative scoring. The teacher maintains complete control by making professionally justifiable scoring decisions after the student responses have been collected and seen. Experienced teachers are thoroughly familiar with such practices.

The reconsiderative method is partially subjective; the teacher is actively making decisions based upon his/her knowledge of the subject matter. However, the response type generally has the characteristics of objective assessments. When such scoring could be done only with manual methods, no special name was necessary. It was simply called fill-in-the-blank or completion testing. With machine-assistance now available, the use of the term reconsiderative test scoring appears to be appropriate.

A most simple explanation of machine-assistance for reconsiderative scoring starts by visualizing an item analysis displayed on a microcomputer monitor screen. (See Figure 1.) Each line gives the frequency tabulation and percentage of how many students selected response A, B, C, etc. By moving

Response	Points	Frequency	Percentage
A		14	11.7
B		35	29.2
C		7	5.8
D	2	43	35.8
E		21	17.5
TOTALS		120	100.0

Figure 1: Frequency tabulation for reconsidering five responses of a traditional multiple choice question.

the screen cursor (place designator) up and down the lines in an additional column called "Points", the teacher can enter from the keyboard the number of points each response should receive. For example, the B's could receive 3 points, the D's receive 1 point, and all others receive zero points. Then the computer re-scores every student's response and allocates the points that were designated by the teacher.

The concept is quite simple, but its usefulness in the context of multiple choice questions appears, at first glance, to be minimal. However, reconsiderative scoring is founded upon fill-in-the-blank (completion-style) questions, not on multiple choice. An item analysis tabulation for an un-cued (no options presented) completion question could look like Figure 2. (The free-response style question is also provided in the caption of the figure.) Such an item analysis has two major differences compared with the previous multiple choice example.

Code	Answer	Pts.	Sub.	Freq.	Percent
000	(Blank)	0	0	2	2.4
186	Canada	2	7	1	1.2
307	England	1	7	6	7.1
325	France	0	0	1	1.2
328	Great Britain	0	0	4	4.7
412	Ireland	0	0	3	3.5
537	United Kingdom	2	7	67	78.8
562	Yugoslavia	0	0	1	1.2
	TOTALS			85	100.0

Figure 2: Example of a microcomputer display for reconsiderative scoring. A class of eighty-five students could be asked a free-response question: "Elizabeth II is the queen of what country?", with the following on-screen item analysis. The point values are designated by the teacher when moving the cursor up and down in the "Points" column.

1. The un-cued responses are not limited to a selection of five choices, and
2. The actual word responses appear on the computer screen.

When dealing with information as shown in Figure 2, the teacher can use his/her content knowledge in a reconsiderative way to evaluate each response. After reading the responses, the teacher designates the points earned by each. Then the teacher presses a key and the micro-computer does all the re-scoring.

B. Computer-Assisted Scoring of Fill-In-The-Blank Responses

But one might ask "How can we generate with existing technology an item analysis showing the actual word responses to completion (fill-in-the-blank) questions?" Such capabilities have been available since 1983 and are fully described in a 1987 book by Anderson. (This book is on ERIC microfiche; see item A1 in the Annotated Bibliography.) The software program is called the MDT Educational Testing System.

The MDT multi-digit method is essentially a fill-in-the-blank (completion-style) assessment that can be scored by machine. Appendix A provides examples, explanations and comments about prior research. Of special importance in Appendix A is the discussion of the improved feedback to teachers and students, including the actual words used to respond to each question. Variations of this relatively new testing method have been called by different names: keylist testing, long-menu questions, un-cued responses, answer bank and multi-digit testing.

> [NOTE: Because of the importance of this method for the development of reconsiderative scoring, readers are encouraged to read Appendix A before proceeding.]

Three recent independent research studies from Quebec (Brailovsky, Bordage, Allen and Dumont, 1988), Pennsylvania (Veloski, Rabinowitz and Robeson, 1988) and Illinois (Anderson, 1988) support the use of the keylist/multi-digit/answer bank method. The research results indicate the following:

1. Strengthened face validity (Veloski et al., 1988),
2. Appropriate use in assessment of student's diagnostic skills (Brailovsky et al., 1988),
3. Appropriate use with certain higher level problem-solving skills that cannot be tested by multiple choice questions (Veloski et al., 1988),
4. Improved identification of marginal examinees (Veloski et al., 1988),
5. General (but sometimes reluctant) acceptance by students (Anderson, 1988),
6. Evaluative power perceived by students to be equal to that of fill-in-the-blank questions (Anderson, 1988),
7. Operational advantages over manually-scored handwritten short answer questions (Brailovsky et al., 1988), and
8. Equal reliability [of scanning] and economy when compared with multiple choice questions (Veloski et al., 1988).

The MDT innovation of machine-scored multi-digit testing is easily understood and can be used by literally hundreds of thousands of educators at all grade levels and in all academic areas. Early use of the MDT innovation has occurred mainly in universities and medical schools. Subject areas range from art appreciation, military science and geography to mathematics and histology. Two national medical

accreditation examinations (Family Practice and Ophthalmology) already are conducting pilot assessments with multi-digit long-list responses. High schools and junior highs can definitely benefit. For example, sophomore geometry final examinations in one high school have used the MDT format for three years. Elementary education may use it in modified forms, such as with two-digit responses.

The lists (answer banks) and question banks could be prepared by individual users or by specialists in each field. These instructional materials could be disseminated to all and modified by anyone. Most of the terms for the lists come from indexes of textbooks, and the "MDT List Maker" software produces the appropriate data file. Participation by academic professional societies, textbook publishers, or sponsors of modest grants will greatly assist the preparation of these learning materials.

Onto these capabilities of MDT multi-digit testing, the new features of reconsiderative scoring have been added.

II. Initial Experiences with Computer-Assisted Reconsiderative Scoring

Computer-assisted reconsiderative scoring was first available in an experimental software package called "RECON" in the Fall Semester of 1989. The author (Anderson) of this conference paper was the developer and first person to use RECON. One copy was sent to William Craig, Director of Testing of the Byron (Illinois) Community Unit School District #226. Changes were made and the functional beta version of the software was distributed in mid-March 1990 to five additional experienced users of the keylist multi-digit method and software. Therefore, only Anderson's experiences in university classes and Craig's secondary school efforts are available to report at this time.

A. Synonyms and Multiple Correct Responses

As described above, reconsiderative scoring allows for more than one correct response. Prior to having the RECON software, the recommended procedure with the multi-digit method was primarily to avoid such questions. However, occasionally a manual review of the printed item analysis tabulation (showing word responses for each question) might reveal that two or more terms on the answer bank long-list could be considered correct. Then the teacher could look at a printed array (as shown in Anderson, 1987, p.99) to see for each student the full listing of the MDT code numbers in columns for each test question. By finding the MDT numbers of the additional correct answers, the points could be manually added to the student scores.

Furthermore, because of this early limitation, the preparation of the answer bank lists either required extra caution to avoid synonyms or the use of duplicate code numbers for synonyms. Because of these difficulties, the use of questions with multiple correct responses was undesirable before the development of RECON.

Although the removal of those difficulties was a major improvement, an even greater gain is the increased flexibility to generate truly challenging questions. Now, with reconsiderative scoring, synonyms are desirable in the answer banks. Diverse responses that communicate the same message of students' understanding can all receive the same allocation of points. Questions no longer need to be worded with so much precision to guide the students toward a preferred response. For example, one MDT long-list question in a pre-RECON World Geography test asked the student "Give the name of the religion that has a pilgrimage known as the Hajj, but do not give the name of a believer of that religion." "Islam" is the correct answer, and the response "Moslem," a follower of the Islamic religion, was not the desired

answer. To overcome this problem, reconsiderative scoring permits the teacher to decide after the responses are collected which answers will receive full (or partial) credit.

Crai (letter, 26 March 1990) illustrates the use of synonyms with a vocabulary item from the Byron high school sections of English I and American Literature: "The _(blank)_ of a caterpillar into a butterfly is a wondrous process." Both 'transformation' and 'metamorphosis' are correct responses.

B. Partial Credit

The Islam/Moslem example above could easily result in partial credit being awarded if the teacher so decides. In other words, "close" can finally count for something. The teacher looks at the responses and then gives partial credit in the same manner that full credit was given. This capability is excellent to stimulate class discussion to distinguish between fully and partially correct answers. As an illustration, when Figure 2 was shown recently to a professor of education, the discussion quickly focused on the partial credit for the response "England." Two geography professors debated with him that although Elizabeth II is the queen of England, the England of today is only one part of the country called the United Kingdom, as is Wales, etc. Because knowledge and thinking are more than black/white and right/wrong, the ability to easily award partial credit with machine assistance should fill a very useful niche in student assessments at all grade levels.

C. Numeric Responses

Three-digit numeric responses (not code numbers from answer banks) have been used since the beginning of multi-digit testing in 1983. For example, 100 is the correct answer to "How many senators are there in the U.S. Congress?" (See Anderson, 1987, pp. 11, 32-33, 64-66, 82.) Care was needed to avoid questions with multiple correct answers. Now, greater flexibility in representation and range of numeric responses can be permitted for reconsiderative scoring (as in Figure 3). A precise number may or may not be required by the teacher, depending on the grade level and objectives of the class. This

Code	Response	Pts.	Sub.	Freq.	Percent
004		0	0	2	6.5
031		0	0	1	31.0
042		1	4	4	12.9
043		2	4	13	41.9
044		2	4	7	22.6
045		1	4	2	6.5
114		0	0	1	3.2
430		0	0	1	3.2
	TOTALS			31	100.0

Figure 3: Reconsiderative scoring of numeric responses. A science laboratory exercise about measurement could ask the following question: "To the nearest whole gram, what is the weight of the yellow precipitate in experiment J?"

flexibility is extremely useful for questions such as "How many millions of people live in the USA in the late 1980s? (For example, if your answer is 83 million, you should encode 083.)" For most classes, an answer close to 245 is correct. Plus or minus 3 million could get full credit. An answer of 235 or 255, although incorrect, is certainly worth more than 172 or 426. The teacher decides.

III. The RECON Contribution to Multiple Choice Questions

A. Toward the Demise of TRADITIONAL "multiple choice"

Anderson has never been totally against multiple choice; he always envisioned the MDT answer bank method existing side by side with multiple choice to fulfill different objectives. But recent experiences with the new reconsiderative software capabilities have revealed how some limitations of traditional multiple choice questions can be overcome by the capabilities of RECON reconsiderative scoring. The advent of reconsiderative capabilities described below could be a serious challenge to the traditional usage of multiple choice questions.

The term "multiple choice" appears to be a misnomer in its traditional usage where the student is instructed to make a single choice of one response out of five (ABCDE) alternatives. That traditional method should really be called "single choice from limited pool" questions. Note this important difference: A true "multiple" would be to select any combination of the letters, such as CE or ACDE, from the same limited pool. There are exactly thirty-two (32) possible combinations (see Figure 4) ranging from none

101 A	106 AB	116 ABC	126 ABCD
102 B	107 AC	117 ABD	127 ABCE
103 C	108 AD	118 ABE	128 ABDE
104 D	109 AE	119 ACD	129 ACDE
105 E	110 BC	120 ACE	130 BCDE
---	111 BD	121 ADE	---
	112 BE	122 BCD	131 ABCDE
	113 CD	123 BCE	(all)
	114 CE	124 BDE	132 (none
	115 DE	125 CDE	of them)
	---	---	---

Figure 4: Thirty-two possible combinations of five letters, each with an MDT multi-digit number. Any question with up to five alternatives labeled A, B, C, D, and E could be used with this special MDT list for "multi-letter" responses. For example: "Which of the following characteristics is/are commonly associated with [whatever topic or situation the teacher chooses to present]: A) ...[word, phrase, sentence or even paragraph]... B) ... C) ... D) ... E) ...

to all five letters. [Note: 32 = 2 raised to the fifth exponential power.] With four letters, there are 16 combinations; three letters yield 8 combinations. For lack of a better name, this could be called either the "multi-letter" format or the "power" format of responses. When each of the combinations is assigned a three-digit number, each combination becomes eligible for designation by students on an MDT multi-digit answer form for machine scoring. These responses would be scored using the reconsiderative methods, as discussed below.

With this "multi-letter" format, questions with five statements can become more challenging. The process of elimination is no longer such a major factor. "Multiple-guess" is no longer one-out-of-five (20%); blind guessing has an almost negligible one-out-of-thirty-two (3.125%) probability of picking the single best answer.

Some educators might argue that this multi-letter approach has reduced the question to merely five True/False statements to be individually accepted or rejected in the response. But close examination reveals that the same criticism also could be made of the single letter traditional version of multiple choice questions. The advocates of multiple choice questions have long felt that the use of five statements together is generally superior to five separate statements.

B. Encouragement for Use of TRUE "Multiple" Choice

One of the biggest unanticipated findings for Anderson was that reconsiderative scoring can make a major contribution to true multiple choice testing. The ease of using the RECON software brings renewed vigor to the use of questions with a selection of four or five responses.

With the student responses shown on the computer screen, the teacher proceeds to allocate an appropriate number of points to each of the multi-letter responses. If response "CE" is the best answer and is worth three points (and "ABD" is totally wrong), then what is the point value of "DE", or "BCDE"? Values of 3, 2, 1 or no points could be assigned. The teacher decides, making a professionally qualified decision with regard to the nature of the five statements.

With the multi-letter format of responses, students can reveal more of their thinking. For example, if one of the five offered statements is patently incorrect, any student who includes that letter is indicating serious deficiencies. Likewise, any definitely correct options should not be excluded from the multi-letter response. Also revealing is the inclusion or exclusion of the other offered statements, either individually or in association with others.

To facilitate this point allocation process, tables for manual assistance are being prepared to specify probable point values to cover most situations. The teacher's subjectivity is blended with professional competence to make the decisions. For example, the teacher could designate that alternative statement "E" has high weight for inclusion while statement "B" has high weight for exclusion in the multi-letter responses that receive the highest points.

Although these computer-aided reconsiderative procedures are much faster than manual scoring of multi-letter responses, they do require more time to score than do the traditional single choice questions. However, that situation may change. The multi-letter responses are highly compatible with computer-assisted item banks and test maker software. There are at least three options for developing these automated capabilities. First, for each multi-letter ("power") question, the item bank can easily store the information about which answers, i.e., combinations of letters, are to receive full credit and which merit partial credit. Second, the computer could be programmed to identify which table of partial points it would use with each item in the question bank. Third, each of the five statements could be assigned a value (based on difficulty or importance) and a computational algorithm could calculate the appropriate point value for any combination of letters.

Some test developers (who write items, not software) are already engaged with true multiple choice questions. In the 1980s there has been an increase in the use of multiple choice questions that ask the student to "mark all that are correct". The annual state assessments in Illinois public schools have

incorporated this format. This multiple mark approach has required the use of OMR optical mark readers (scanners) that allow more than one mark in an answer grid. That requirement is highly contrary to the capabilities of high quality scanners to distinguish between light marks and poor erasures. In other words, the multiple mark educational measurement method is basically at odds with the hardware capabilities. The reconsiderative method as described above can provide TRUE multiple choice capabilities as good as or actually better than existing solutions that use multiple marks, unique wordings of the five choices, multiple answer keys or more costly scanners.

C. Summary Comments about Multiple Choice Questions

In recent years strong statements have been made both in attack and defense of multiple choice testing in education. Much of the conflict could be attributed to the fact that the multiple choice format has been the only pencil-and-paper, machine-scored method of educational measurement capable of economically scoring millions of student responses while generating useful statistics. Multiple choice has not been a dead end street; it has been worthy of defense. But multiple choice is restrictive and contrived. It is not "natural;" the natural decisions in daily life are not based upon five choices, of which only one can correct.

The origins of multiple choice testing (summarized in Anderson, 1988) were in the early Twentieth Century. Since then the method has been enhanced with research, statistics, machine scoring and some interesting applications, the best of which is probably adaptive assessment with each student at a computer terminal. The unquestionable greatest strength of multiple choice testing is its ease of scoring, whether by hand or by optical mark readers. But the method is still primarily based on five choices that are more difficult to devise than the question itself with its correct answer.

Enter the capabilities of reconsiderative scoring and multi-digit/answer bank responses. The answer bank method is founded upon fill-in-the-blank/completion testing that is far older than the multiple choice method. And it can be scored by a machine! Likewise, the reconsiderative method is very old in concept but very new in machine-assisted applications. For example, the "multi-letter" variation discussed above was not thought of until the beginning of 1990. And major additional enhancements can be made, as discussed in Section V.

Simply stated, the traditional multiple choice format is no longer the only serious contender for machine scoring of student responses, whether in small classes or in nationwide assessments. This is not a statement against multiple choice testing, which undoubtedly has a continuing niche to fill. This is a statement in favor of better education through a greater variety of more academically rigorous assessment methods.

IV. Student Opinions

Student attitudes about MDT multi-digit testing have been reported in earlier studies and are cited in Appendix A (Anderson, 1988, is the most recent and best summary of those findings). Anderson's classes that have been exposed to reconsiderative scoring total approximately 250 students. The students have each taken two or three examinations and have not indicated any difficulty in understanding how to respond to the new formats of questions. Nor have they considered the method unfair or inappropriate. Essentially, the students have attitudes that are neutral to favorable. They say that the MDT method plus reconsiderative scoring makes tests more difficult, but fair. Further research with a short questionnaire of student opinions is planned.

V. Additional Capabilities

[NOTE: A detailed discussion of these capabilities is scheduled as a keynote address by Anderson at the Second International Computer-Managed Learning Conference on 16-18 May 1990 in Edmonton, Alberta, Canada.]

The MDT and RECON innovations are only the tip of an iceberg of software to enhance educational measurement. Some of the items listed below could be developed quite quickly as additions to the existing software; others will require time, funding and perspiration. All can be accomplished; we do not know if and when all should be developed.

A. Multiple Responses to One Question

As shown in Figure 5, one question can require several responses that could be given in any order. That example also illustrates use with higher order questions.

Code	Response	Pts.	Sub.	Freq.	Percent
017	Actinomycosis	0	0	18	22.5
102	Bacterial Meningitis	3	9	56	70.0
103	Bacterial Meningo-encephalomyelitis	0	0	21	26.3
806	Rabies	3	9	77	96.3
907	Tetanus	0	0	5	6.3
916	Thromboembolic Meningo-encephalitis	3	9	62	77.5
	TOTALS			560	700.0

Figure 5: Complex medical diagnosis question: Questions 2-8: Give seven differential diagnoses for the following case. Data: Hereford, 650 lbs., feedlot steer, vaccinated (IBR/BVD/PI3). Symptoms: Sudden onset of blindness, tremors, frothy salivation, opisthotonos, gets better, then gets worse. (Class size is 80 students, so 560 responses (7 x 80) are scored and tabulated.)

Questions requiring multiple responses to a single question can already be used with the existing software, but care must be taken to verify manually on one printed report that the same answer was not given more than once by each student. For example, 'Rabies' could be used only once, not seven times by the same student. That data check can be incorporated into the software. Also, in future versions the program will tally the multiple responses into one item analysis, hence the total being 300 percent for three responses or 700 percent for seven responses.

This multiple response capability is extremely powerful because questions can be phrased in so many challenging ways. For example, some correct answers could be eliminated while also guiding the students to understand the nature of the question, as in: "Name two South American countries (other than Brazil) that have substantial areas of tropical rainforest." A multiple response question from the Byron school district (Craig, letter, 26 March 1990) was "List three of the five characteristics of mammals."

B. Enhanced Numeric Responses

Response grids for longer numeric answers plus a machine readable decimal point are extremely important for mathematics and science education. An example is in Figure 6. The solution for marking and scanning will be compatible with the MDT and RECON software and will utilize existing models of sheet and card readers.

Response	Points in Sub. S1	S2	S3	Freq.	Percent
(Blank)				564	0.400
.013		1		3,447	2.700
.04				69	0.001
.13		1		7,431	5.820
.23				36	0.000
.3				843	0.660
.40				16	0.000
.85				73	0.001
1.2	1			4,087	3.201
1.3	1	1		104,151	81.569
1.4	1			3,123	2.516
*3				87	0.001
4				21	0.000
8.5				18	0.000
13.		1		3,263	2.556
85.				6	0.000
1 3	1	1		21	0.000
246				1	0.000
13000		1		427	0.334
TOTALS				127,684	100.000

Figure 6: Simple numeric response in a standardized test. The wording of the question could be a story problem in which the student is to add 0.8 plus 0.5 to test addition of decimals. Note that the ability to add and the ability to place the decimal point are two separate tasks, for which students receive points under columns S2 and S1, respectively.

C. Multiple Steps in Responses

Truly challenging questions are frequently complex, involving a series of steps (intermediate products) to reach a final answer. The multi-digit responses (as numbers or as words) could be the input to equation solvers and concept mapping software. A reconsiderative capability could be used to assess multi-term and multi-step responses that reveal student thinking. This capability is BIG and could utilize artificial intelligence to assist teachers!

D. Graphical Responses

Scanning and reconsiderative grading of student graphical responses will open new dimensions for assessment. This capability will require different scanners, but the technology is already available and will include at least computer-recognition of handwritten numerals.

E. Improved Feedback

Better feedback to teachers, students and parents can utilize the already available "vocabulary" for correct and actual responses in the answer bank lists used with the MDT software. When coupled with improved statistical analyses (because guessing is almost eliminated), the improved feedback can assist adaptive instruction for remedial, regular and advanced study.

F. Links with Computer Managed Learning and Databases

Existing question banks (test generators) and administrative databases could be modified to utilize the completion-style responses of keylist/answer bank assessments. Data exchanges via networks, workstations and large file servers on micro, mini and mainframe computers will place major assessment and instructional power at the fingertips of teachers, and all steps can be very "user friendly."

And more and more possible enhancements are yet to come. Anderson believes that he has only scratched the surface. Significant new applications and capabilities have become evident to him every semester for the past seven years. When more researchers with diverse backgrounds become involved, the pace of development is expected to accelerate.

VI. Operational and Financial Issues

The RECON and MDT software packages currently use common MS-DOS microcomputers. The more powerful microcomputers with 386 or 486 processors are recommended for professors with large classes. Reconsiderative scoring requires a fantastic number of calculations and data checks, so power and speed are highly desirable. Finally, here is computer software that clearly justifies the acquisition of truly powerful microcomputers for educators.

Standard answer form readers for sheets or cards are compatible. Almost any optical mark reader (OMR = scanner for test scoring) can be used, ranging from inexpensive ($500.00) manual-feed card readers to high-priced, high-speed sheet readers. The answer forms cost the same as those purchased for multiple choice tests. The entire system is designed to be available to and used by individual teachers or by centralized school offices for measurement and evaluation.

Although developed for classroom assessments conducted by individual teachers, the reconsiderative method also could be applied to standardized, norm referenced tests. The ease of

generating test items with the MDT answer bank method will reduce the need for test security. In an educational environment where tests do at least influence the curriculum, the ability to freely disseminate question banks as well as answer banks should have a favorable impact.

Interested potential users should direct their inquiries to the various suppliers of scanners, answer forms, software and computers to obtain the most current information on supply, price and compatibility.

VII. Educational Implications and Research Issues

The advent of reconsiderative scoring (as concept and as functional software) should stimulate both research in the concept and the reporting of applied experiences with the software. Numerous topics of inquiry are evident, not the least of which are the relationships to test theory.

A. Relevance to Assessment Theory

A fundamental question is this: "How do the MDT and RECON capabilities support or conflict with theories of education?"

Robert J. Mislevy's work on "Foundations of a New Test Theory" (1990) includes the following insights: "Tomorrow's tests must present tasks that learners in the different states [of competency] are likely to carry out in observably different ways. We cannot limit our interest to the correct response, but must also consider factors such as speed, intermediate products, and incorrect responses. We must also examine the patterns of similarity or dissimilarity across tasks that probe knowledge structures or problem-solving techniques. The new test theory must provide models that can express these patterns."

At least at first glance, the characteristics of multi-digit answer bank questions plus the reconsiderative scoring capabilities appear to be supportive of Mislevy's desired models and methods. This support could be especially important when working with large numbers of students that make at least some machine assistance an economic necessity.

Numerous theoretical concerns need to be addressed in future presentations. The paragraphs below highlight a few of the educational implications and topics appropriate for further research.

B. Reduction of Bias

Because each item analysis is a tabulation of the student responses, the computer makes certain that each and every student with response "XXX" receives precisely the same number of designated points. This eliminates inconsistencies and bias that can occur in manual scoring of completion questions when the teacher spends hours to go from the top to the bottom of a stack of examinations.

C. Ease of Writing Questions

The educational benefits of the reconsiderative multi-digit format include the ease of writing questions that do not need four wrong but plausible foils. When used with diagrams such as those in Figure 7, hundreds of questions are easily generated.

D. Increase in Academic Rigor

A third major benefit is the increased academic rigor over similar multiple choice questions. If a discrete answer to a question is known, calculated or derived by thought, then the desired response can be located easily in the alphabetized list. But if not known or derived, the correct response is only one out of many, and not one out of five. Guessing is senseless; the process-of-elimination is applicable only if the student has studied sufficiently to have the appropriate vocabulary and concepts in mind.

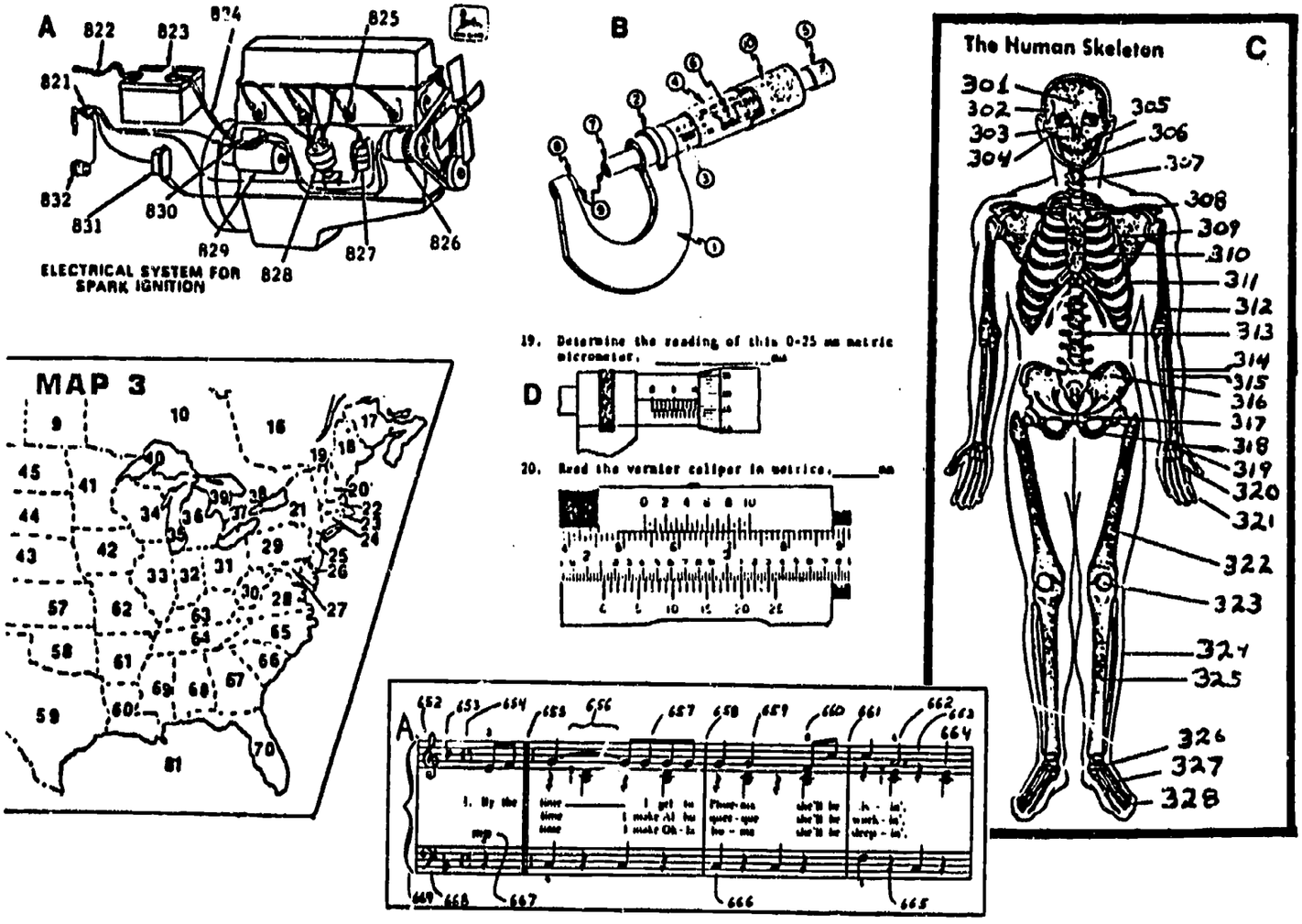

Figure 7: Diagrams compatible with MDT-style questions.

E. Application for Instruction as Well as Assessment

The MDT format plus reconsiderative scoring can be used for tests, exercises and homework, whether for grades or as learning experiences. "The promise held forth by the MDT software and the RECON module is not merely on the assessment side, it seems to me, but on the instructional side as well. Using this software to assess our students' mastery of local goals and objectives imparts an intensified classroom focus on those goals and objectives" (Craig, letter, 26 March 1990).

F. Inclusion of Subjective Input

The RECON method allows teachers to have subjective input where appropriate. Reconsiderative scoring could stimulate greater usage of higher order questions.

G. Ease of Understanding

All indications are that machine-assisted reconsiderative scoring can be easily understood and utilized by teachers and students in upper elementary, secondary and post-secondary education, including vocational and professional training. Being very natural, like manual reconsiderative scoring, the method could have wide acceptance, as discussed below.

VIII. Comments on Adoption and Dissemination

A. The Underlying Conceptual Model

Being so simple, the RECON capability almost defies being called an innovation. But it is a classic example of how innovative ways can ease the burden of age-old tasks. Teachers have been doing reconsiderative grading of responses one-by-one for centuries. They undoubtedly understand the task. Now, with computer speed and ease to let teachers virtually see all student answer sheets at the same time in tabulated form, teachers do faster and better the same job that they already know and do so well. Info-World Magazine (13 February 1989, page 1) discusses three factors needed for a truly user-friendly software interface: the screen, the command structure, and "even more important, ... the underlying conceptual model of the software." The spreadsheet concept made VisiCalc and Lotus 1-2-3 great programs for this reason: "experienced financial analysts can pick up the program in 15 minutes, because they understand the underlying task so well." Likewise, teachers already understand the task of reconsiderative scoring. RECON is software that teachers can readily use.

B. The Problem of Inertia

On the other hand, the biggest threat to the practical application of the MDT and RECON innovations is the inertia in all levels of America's schools. Instead of lamenting that stumbling block to all innovations, we can explore three ways to overcome the inertia that hinders the adoption of computer-assisted reconsiderative scoring.

1. The first way is through Anderson whose company (MDT Corp.) can influence the availability of the MDT/RECON software. The software already exists with commercial quality and includes standard capabilities for multiple choice and criterion referenced testing. Anderson's difficult double role as a businessman as well as an academic innovator is not the topic of this academic paper, but that role is an important factor for keeping the software affordable for the maximum number of schools.

2. The second way focuses on getting microcomputers and essential software into the hands of teachers. Anderson believes that fully integrated reconsiderative scoring, as described in Section V above, can become for teachers what word processors are for secretaries and writers, spreadsheets are for accountants, and databases are for managers. And the reason is this: If a teacher and his/her students are to receive the discussed benefits, the teacher (or assistant) is actually required to look at the microcomputer screen and use arrow and number keys to interact with the student responses. The only alternative method for reconsiderative scoring is slow manual grading. And that manual scoring looses the benefits of computer-generated feedback to improve student learning and question quality. Typewriters, pocket calculators, card indexes, and even stand-alone multiple-choice grading machines have either allowed many teachers to avoid using computers or have allowed some education administrators to say that students, not teachers, should get the microcomputers. Instead, all teachers should have microcomputers and essential software (including reconsiderative scoring) for their daily tasks.

3. The third way to overcome inertia in education is strongly influenced by organizations and corporations with a vested interest in the advancement of computer-assisted education. They can have major impact upon what actually becomes known to and accepted by educators. However, they can also compound the problem of inertia when their self-proclaimed leadership roles as "champions of education" are merely lip-service while their true roles are those of followers of profit, as discussed below.

IX. Initial Experiences with External Sponsors

The inertia in the "champions of education" is clearly shown in the initial experiences at the corporate level concerning reconsiderative scoring. MDT Corporation has made disclosures of the RECON concept and capabilities to numerous entities in the past year. The objective of the disclosures was to find a sponsor, partner, or advocate to assist in the development of reconsiderative scoring methods. These entities included IBM, Zenith, National Computer Systems (NCS), ScanTron, HEI Scanning Systems, ETS, ACT, several major textbook publishers, US Department of Education, state edueation agencies in Illinois, and major software developers like MicroSoft and Lotus. To date, every one has adopted either a "We don't do that" attitude or a low-risk, "market-driven", "not-invented-here", wait-and-see attitude to determine where profits (not educational improvements) can be found. This is not a criticism of these entities; it is only a clear statement of the obstacles that confront any innovation in education.

What John Roach (Chairman of Tandy Corporation) said in 1984 about the microcomputer industry is quite applicable to the issues of educational technology: "We're in an industry where promotion is more important than the technology." Sad, but true: Without recognition and promotion, innovations cannot attain practicality.

The bright side is that if educators can show that a sufficient market does exist, one or many of the corporate and not-for-profit entities would gladly and efficiently carry the banner of educational innovation and even reform. However, inertia and conservatism among educators and educational measurement specialists could lead to very slow or no acceptance of computer-assisted reconsiderative scoring. To paraphrase some comments from the above named entities, "Teachers don't want this stuff. They don't know how to use it, and many are afraid of computers. This looks like more work, not less. Besides, students don't want tougher tests, and parents don't like being shown how little their kids know. It won't sell easily, and we are not in the education reform business. But we'll be glad to work with you when the market is evident."

X. Conclusion

Although the initial in-course experiences with reconsiderative scoring are quite favorable and encouraging, no conclusions can be made on the merits of the new method until more users have results to report. The willingness of research-minded educators to examine objectively the issues associated with reconsiderative scoring is crucial. The author firmly believes that the method will withstand the most vigorous scrutiny. Furthermore, with additional users, still more innovative ways to enhance the initial capabilities will be discovered. This present (April 1990) presentation is intended to introduce and stimulate discussion about one additional measurement technique made possible by educational technology.

APPENDIX A: Background to MDT Multi-Digit Testing

I. Origin

In 1982 a geography professor, Dr. Paul S. Anderson, returned to the USA after fourteen years in Australia and Latin America. For the first time he was faced with large classes needing computer-assisted testing. The unwelcome necessity to use multiple choice tests prompted him to seek a more rigorous alternative. In four months he conceptualized and made operational a method of computer-assisted scoring of fill-in-the-blank (completion-style) questions. Anderson's geography training in cartography, remote sensing, computer analysis, and regional studies gave him, respectively, the necessary background to print answer sheets, understand electronic scanners, appreciate computer power and utilize alphabetized long lists of terms, as in an atlas gazetteer. What emerged has far exceeded his initial expectations and has led him on a lateral career path to explore and develop the potential of his initial innovation. After seven years of extra hours, heavy investment of personal funds, and even times of anguish, Anderson has produced two fully operational innovations for educational measurement, and further innovations can be clearly seen.

The innovation for machine scoring of fill-in-the-blank questions that Anderson first used in early 1983 is described in his book, The MDT Innovation (1987, available in all libraries with ERIC microfiche). The 200-page book discusses the method's origins, initial usage, applications and educational implications, including financial savings and relevance to higher order learning.

II. Method

The MDT innovation is so straight-forward that it can be explained in a single diagram (Figure 8) found on the book's cover.

A. Completion-style questions are asked in a wide variety of ways, including some that require numeric responses.

B. With an answer in mind, the student locates his/her desired response on an alphabetized long-list "answer bank" appropriate for the subject area. The list is quite important, but it can be made easily by any teacher or shared by teachers of similar courses.

C. The three-digit code number of the response is marked on an answer form that is read by standard optical mark readers of sheets or cards.

D. Microcomputer software processes the student responses, issues their scores and prints distinctly useful reports.

III. Distinguishing Differences from Prior Efforts

With such a simple and natural methodology, was nothing like this ever tried previously? Chapter 8 in The MDT Innovation (Anderson, 1987) reviews prior efforts fortunately not known to Anderson when he began. Anderson's work is distinctive in two crucial ways.

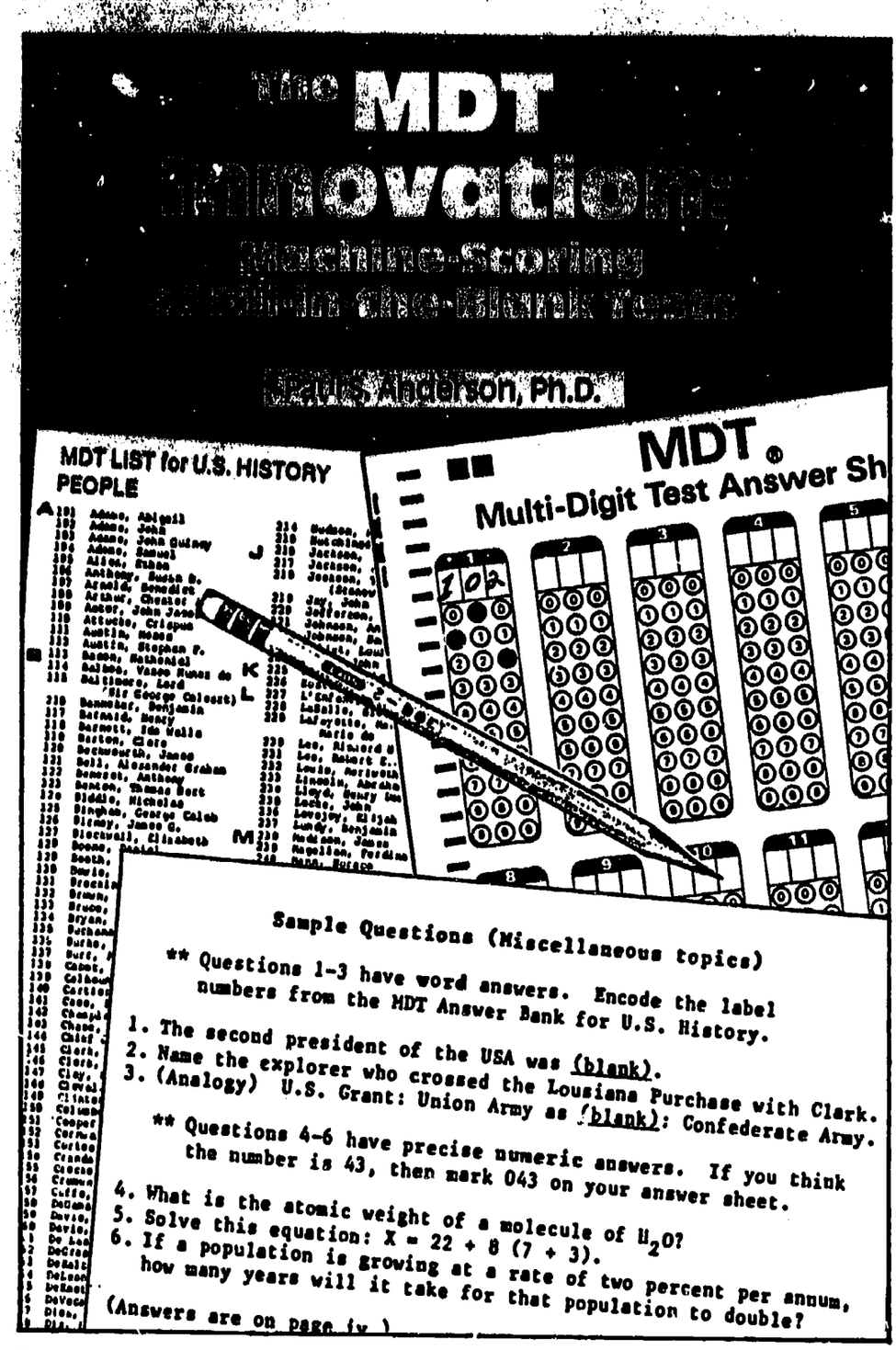

Figure 8: Examples of MDT-style questions, lists and answer forms.

First, no previous researcher had significantly gone beyond one hundred responses, that is, responses with two digits. Anderson began his work with three-digit numbers that allow up to 1000 terms and concepts in the answer bank. Each list is intentionally long to nearly eliminate the "I'll recognize the answer if I see it" effect. Short lists encourage and permit searching that waste time when students try to recognize answers. This is a fundamental difference from "matching" tests. List length should be in consideration of the grade level of the students, but totally at the discretion of the test writer. For university students, Anderson recommends lists of over 500 terms. Furthermore, he uses response number 999 to mean "no reasonable answer is on this list." When responding with 999, the student also writes the correct word response in a blank space on the answer sheet. The evident strength of the MDT method is the high degree to which it emulates completion-style (fill-in-the-blank) responses, of which the academic rigor and validity are well established.

Second, earlier researchers did not develop any computer program to do the scoring. Consequently, Anderson was the first to observe and use the tremendous time-saving advantages of these computer-scored completion tests. Even more important, he recognized and made operational the capability to generate some truly distinctive feedback for teachers and students. The most innovative feedback involves the ability to print the actual word responses of the students and of the teacher's answer key. The MDT responses with printed words in the Individual Student Report (Figure 9) give far more assistance to the student than do the multiple choice responses of A B C D or E. The Item Analysis tabulation of responses (Figure 10) gives valuable information to the teacher about student learning.

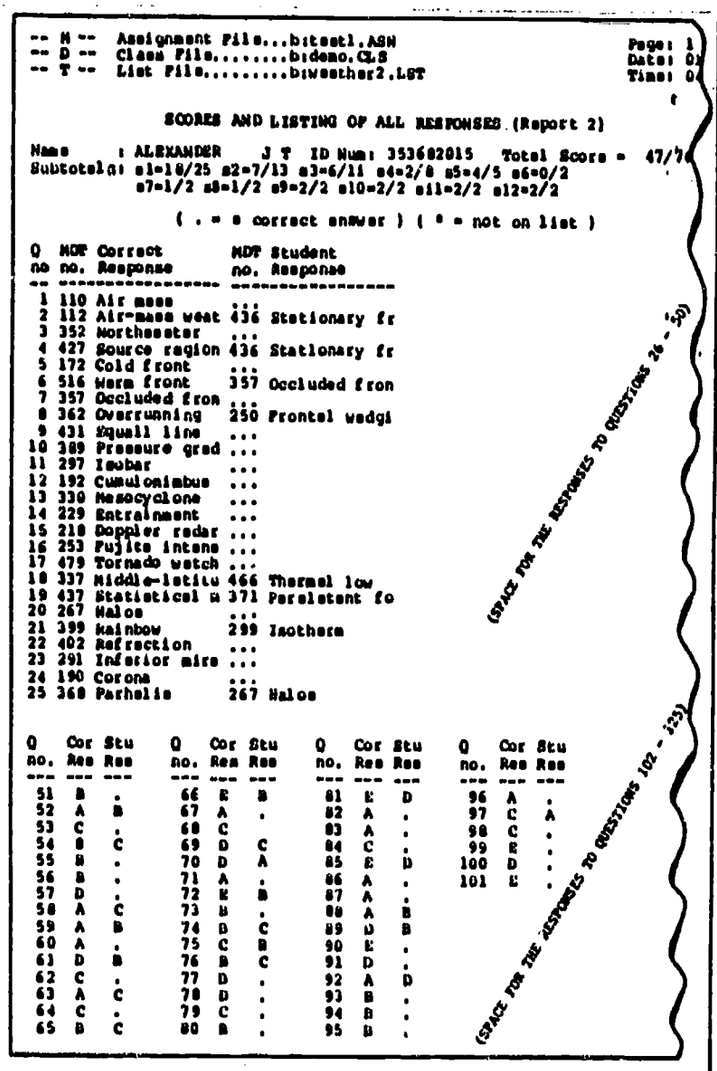

Figure 9: Individual student report with MDT actual word responses.

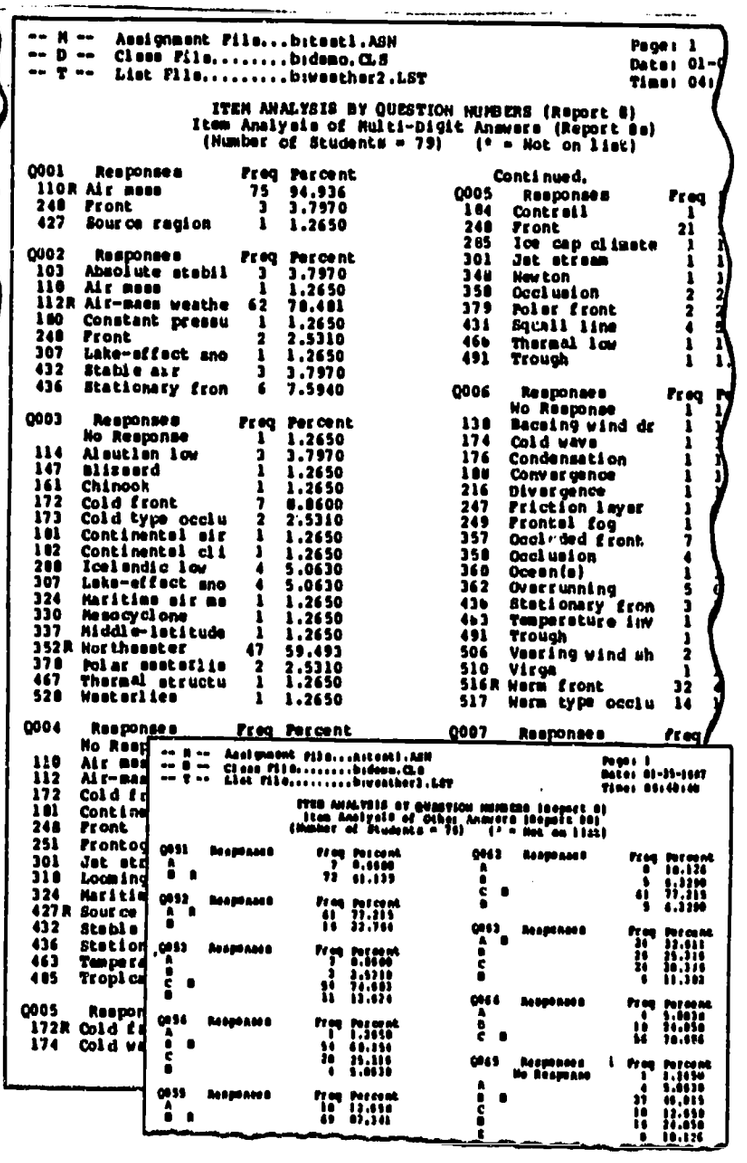

Figure 10: Item analysis with tabulation of student word responses.

Even though he worked with these reports regularly since 1983, Anderson needed nearly four years before the second innovation became evident to him. Only in the 1989-90 academic year has this second innovation become a reality in a software program called RECON for reconsiderative scoring.

BIBLIOGRAPHY OF CITED REFERENCES

(See also the Annotated Bibliography for MDT/RECON.)

Anderson, Paul S. (1987) The MDT Innovation: Machine Scoring of Fill-in-the Blank Tests, Multi-Digit Technologies Corporation, 107 Broadway, Normal, Illinois, 198 pages, 1987. (Also available on three ERIC microfiche: ED 307 287)

Anderson, Paul S. (1988) "An Educology of Testing: American Student Attitudes about Test Formats, with Special Reference to the MDT Multi-Digit Testing Method." International Journal of Educology, Vol. 2, Sydney, Australia, pp. 143-184. 1988. (ERIC EJ 394 496)

Brailovsky, C. A., G. Bordage, T. Allen and H. Dumont (1988). "Writing vs. Coding Diagnostic Impressions in an Examination: Short-Answer vs Long-Menu Responses." Research in Medical Education (RIME) Proceedings, the 27th Annual Conference of the Association of American Medical Colleges, Chicago, 11-17 November 1988, pages 201-206.

Craig, William (1990) (Letter to author, dated 26 March 1990)

Mislevy, Robert J. (1990) Edited excerpt from "Foundations of a New Test Theory," as published in ETS Developments, XXXV/2&3, Fall/Winter 1989-90, pp. 10-11, ETS, Princeton, New Jersey. The excerpt is from a chapter in Test Theory for a New Generation of Tests, edited by Norman Frederiksen, Robert Mislevy and Isaac Bejar, (in press for 1990), Lawrence Erlbaum Associates, Hillsdade, New Jersey.

Veloski, J. Jon, Howard K. Rabinowitz, M.D. and Mary R. Robeson (1988). "Cueing in Multiple Choice Questions: A Reliable, Valid and Economical Solution." Research in Medical Education (RIME) Proceedings, the 27th Annual Conference on Research in Medical Education, Annual Meeting of the Association of American Medical Colleges, Chicago, 11-17 November, 1988, pp. 195-200.

Annotated Bibliography

With Specific Relevance to MDT Multi-Digit Testing and RECON Reconsiderative Scoring

NOTES:

1. This bibliography is divided in several ways. The publications and presentations of Paul S. Anderson, the originator of MDT multi-digit testing and RECON, are presented by types (books, articles, etc.) in chronological order. The works of others are in a separate section in alphabetical order.

2. The designation "ERIC" identifies items available internationally in many libraries via indexes and microfiche. Copies are available from the Educational Resources Information Center, c/o ERIC Document Reproduction Service, Alexandria, VA 22304-6409.

3. All readers are requested to nominate additional items for inclusion in this bibliography. Please send references (and copies, if available) to:

Dr. Paul S. Anderson
Department of Geography-Geology
Illinois State University
Normal, IL 61761

Phone (309) 438-7360 (Office)
(309) 438-7649 (Department)

PUBLISHED ITEMS BY PAUL S. ANDERSON PLUS CO-AUTHORS: (Groups A & B.)

Group A: BOOKS, MONOGRAPHS AND MANUALS:

A1. (1987) The MDT Innovation: Machine Scoring of Fill-in-the Blank Tests, Multi-Digit Technologies Corporation, 107 Broadway, Normal, Illinois, 198 pages, 1987. (ERIC ED 307 287 -- three fiche)

This book ($14.95) is a basic reference. Its sections provide definitions, examples, development background, user notes for teachers and students, sample reports, review of pre-1987 academic references, and discussions of retention of learning, mastery/training, costs, and higher order learning. The book includes much from presentations C1 through C10.

A2. (1987 with revisions) MDT Educational System: User's Guide, with co-author James S. Schoner. Multi-Digit Technologies Corporation, Normal, Illinois, 86 pp., 1987.

This manual is specific to the computer software from MDT Corporation. It also contains examples of reports.

A3. (1989) A Learning Assessment System: Development of Assessment Instruments Plus Scoring and Reporting Procedures, with co-author Larry P. Marsh. Bureau County Learning Assessment Cooperative, with funding from the Illinois State Board of Education, pp. 127, 1989. (ERIC ED 307 855 -- two fiche)

This monograph ($3.50) does not focus on multi-digit testing, but it provides the most "tutorial-like" materials written about the MDT software and interpretation of the reports and statistics from multiple choice questions and criteria referenced testing (CRT).

Group B: ARTICLES AND OTHER PUBLICATIONS:

B1. (1984a) "An Introduction to the Multi-Digit Test (MDT)," Discussion Papers in Geography, No. 2: "Objective Testing in Geography", Old Dominion University, Norfolk, Virginia, pp. IV-1 to IV-18, 1984.

First published item. Revision of presentation C1. Full contents have been incorporated into A1.

B2. (1984b) "Applications of the Multi-Digit Test (MDT) Procedure for Teaching the Geography of Latin America," CLAG Communication, Newsletter No. 50, pp. 2-3, December 1984.

Minor news-note with examples.

B3. (1985) "Comparison of Cognitive Retention from Three Testing Methods: Fill-in-the-Blank, Multiple Choice and the Multi-Digit Test (MDT)," with co-authors Miriam Hill, Shamim Naim and William Walters, Illinois School Research and Development, Journal of the Illinois Association for Supervision and Curriculum Development, Normal, Illinois, Vol. 21, No. 1, pp. 28-37, Winter 1985. (ERIC EJ 315 070)

Fully reprinted in Chapter 8 of A1.

B4. (1987) "Testing---1, 2,...523,...641,...999---Testing: The MDT Multi-Digit Technique Applied to Science Education," with co-author James S. Schoner. Spectrum, Journal of the Illinois Science Teachers Association, Spring/Summer issue, pp. 17-21, 1987.

First major discussion of manual methods for using the MDT multi-digit testing method. Includes examples from cell biology. Includes reproducible (royalty-free) answer sheet for manual scoring.

B5. (1988a) "An Educology of Testing: American Student Attitudes about Test Formats, with Special Reference to the MDT Multi-Digit Testing Method." International Journal of Educology, Vol. 2. Sydney, Australia, pp. 143-184. 1988. (ERIC EJ 394 496)

Major publication of research results from a sample of 144 student. Results indicate strong similarities of the multi-digit and fill-in-the-blank formats. Includes vast majority of C12 and C15.

B6. (1988b) "Uses of MDT Multi-Digit Testing in Geographical Education." Chapter 34 in Rod Gerber and John Lidstone (eds). Developing Skills in Geographical Education. International Geographical Union with Jacaranda Press and Brisbane College of Advanced Education, Australia. pp. 215-221. 1988.

First reference to reconsiderative scoring.

B7. (1988c) "Changing World Patterns of Machine-Scored Objective Testing: The Expected Impact of the Multi-Digit Method." (with co-author Alcyone Saliba). In George Padavil (ed.). Internationalizing Curricula. Occasional Papers of the Mid-west Cooperative and International Education Society Annual Conference, Normal, Il. pp. 177-188. 1988.

Based on presentation C13. Contains typology of educational measurement worldwide. Suggests that multi-digit testing could have significant advantages for international usage.

ITEMS BY OTHER AUTHORS: (Group G.)

G1. Brailovsky, C. A., G. Bordage, T. Allen and H. Dumont (1988). "Writing vs. Coding Diagnostic Impressions in an Examination: Short-Answer vs Long-Menu Responses." Research in Medical Education (RIME) Proceedings, the 27th Annual Conference of the Association of American Medical Colleges, Chicago, 11-17 November 1988, pages 201-206.

Very important and very applicable research. "The results . . . seem to favor the replacement of short-answer questions [completion-style] by long-menu questions [multi-digit style] in the assessment of students' diagnostic skills."

G2. Cook, Desmond L. (1955), "An Investigation of Three Aspects of Free Response and Choice Type Tests at the College Level," Ph.D. Dissertation, University of Iowa, University Microfilms International, Ann Arbor, Michigan.

Comments in A1.

G3. Duchastel, Philippe C. (1981), "Retention of Prose Following Testing with Different Types of Tests," Contemporary Educational Psychology, July 1981, Vol. 6, pp. 217-226.

Comments in A1.

G4. Gay, Lorraine R. (1980), "The Comparative Effects of Multiple Choice Versus Short-Answer Tests on Retention," Journal of Educational Measurement. Spring 1980, Vol. 17, pp. 45-50.

Comments in A1.

G5. Meyer, George. (1934), "An Experimental Study of the Old and New Types of Examination: I. The Effect of the Examination Set on Memory," The Journal of Educational Psychology. December 1934, Vol. 25, pp. 641-661.

Comments in A1.

G6. Sax, G., and L. S. Collett. (1968), "An Empirical Comparison of the Effects of Recall and Multiple-Choice Tests on Student Achievement," Journal of Educational Measurement, Vol. 5, pp. 169-173.

Comments in A1.

G7. Veloski, J. Jon, Howard K. Rabinowitz, M.D. and Mary R. Robeson (1988). "Cueing in Multiple Choice Questions: A Reliable, Valid and Economical Solution." Research in Medical Education (RIME) Proceedings, the 27th Annual Conference on Research in Medical Education, Annual Meeting of the Association of American Medical Colleges, Chicago, 11-17 November, 1988, pp. 195-200.

Highly relevent research. "The results support the feasibility of large group administration of tests conducted in an open-ended format that can be scored by computer. Not only is this format equally reliable and economical when compared with the MCQ [multiple choice questions], but it also provides important advantages that strengthen its face validity. The Un-Q [uncued multi-digit style] format can be used to test either simple recall or certain higher level problem-solving skills that cannot be tested by MCQs. Even more important, the results also suggest that the Un-Q format may be a more effective discriminator of academically marginal examinees."

G8. Ward, William C. (1982), "A Comparison of Free-Response and Multiple Choice Forms of Verbal Aptitude Tests," Applied Psychological Measurement, Vol. 6, No. 1, Winter 1982, pp. 1-11.

Uses term "keylist." Short lists of up to 100 choices. No computer program for scoring. Additional comments in A1.

PRESENTATIONS AND CONFERENCE PAPERS By Paul S. Anderson Plus Co-Authors: (Only items with some reproducible materials.) (Group C.)

C1. (1983) "Introduction to the Large-List 'Multi-Digit Test' Procedure: With Examples From a World Geography Course," West Lakes Regional Conference of the Association of American Geographers, Iowa City, Iowa, October 1983.

See A1 and B1.

C2. (1984a) "Applications of the Multi-Digit-Test (MDT) Procedure for Teaching the Geography of Latin America," Conference of Latin Americanist Geographers (CLAG), Ottawa, Canada, September 1984.

See B2.

C3. (1984b) "The Multi-Digit Test Procedure: Refinements and Preliminary Results," International meeting of the National Council for Geographic Education, Toronto, Canada, October 1984, with co-authors M. Hill, S. Naim and W. Walters.

See A1 and B3.

C4. (1985a) "Laboratory Schools as a Unique Setting for Research: The Experimentation with the Multi-Digit Test (MDT) at Illinois State University High School," with co-author Eileen Kanzler. National Association of Laboratory Schools convention in Denver, Colorado, 26-28 February 1985.

Mainly about the laboratory school setting.

C5. (1985b) "Comparison of Cognitive Achievement in Objective Testing: Multi-Digit and Multiple Choice Tests," with co-author Eileen Kanzler. Conference of the American Educational Research Association (AERA), Chicago, Illinois, 4 April 1985. (ERIC ED 260 131)

Major paper. Much was incorporated into A1. This paper led to the nationwide press coverage in D1.

C6. (1985c) "Applications of the Multi-Digit Test (MDT) in Geography," Demonstration presented to the Conference of the National Council for Geographic Education, Breckenridge, Colorado, 5-9 August 1985.

C7. (1985d) "Innovations in Educational Testing with Optical Mark Readers: Multi-Digit Large-List Tests, Subjective Question Scores and Instant Scoring in the Classroom," Discussion session at the World Conference on Computers in Education, Norfolk, Virginia, 29 July 1985.

C8. (1985e) "Applications of the Multi-Digit Test (MDT) in Science Classes," Workshop at the Illinois Science Teachers Association convention in Normal, Illinois, 4-5 October 1985.

C9. (1986a) "Multi-Digit (MDT) Testing in the Teaching of Criminal Justice Sciences," with co-author Diane Alexander, Academy of Criminal Justice Sciences, Orlando, Florida, March 1986. (ERIC ED 282 936)

C10. (1986b) "Applications of the MDT Multi-Digit Testing Method for Cartographic Information Education," Annual meeting of the North American Cartographic Information Society, Philadelphia, Pennsylvania, 15-18 March 1986.

C11. (1987a) "Answer Banks and Question Banks for Immediate Classroom Use: MDT Materials for Geography and Social Studies," Annual meeting of the Illinois Geographical Society, Elgin, Illinois, 10-11 April 1987.

C12. (1987b) "Student Attitudes about MDT Multi-Digit Testing Analyses from Pioneer Experiences," Annual meeting of the National Council on Measurement in Education, held jointly with the American Educational Research Association, Washington, DC, 19-23 April 1987. (ERIC ED 296 000)

Early results later used in B5 and C15.

C13. (1987a and 1988) "Changing World Patterns of Machine-Scored Objective Testing: The Expected Impact of the Multi-Digit Method," with co-author Alcyone Saliba, Sixth World Congress of Comparative Education, Rio de Janeiro, Brazil, 6-10 July 1987. Also presented to annual conference of the Midwest Comparative and International Education Society, Normal, Illinois, 19-20 February 1988.

See B7.

C14. (1987b) "Advantages of MDT Multi-Digit Testing in Veterinary Medicine Education: Basic Features," (with separated demonstration). Fifth Symposium on Computer Applications in Veterinary Medicine, Urbana, Illinois, 26-29 September 1987.

C15. (1987c) "Comparison of Stude. Attitudes about Seven Formats of Educational Testing, With Emphasis on the MDT Multi-Digit Testing Technique," Annual meeting of the Mid-Western Educational Research Association, Chicago, Illinois, 15-17 October 1987. (ERIC ED 295 999)

Major paper incorporated with C12 into B5.

C16. (1989) "Introduction to Machine-Assisted Reconsiderative Test Scoring: A new method for partial credit and multiple correct responses," Special presentations to medical and science educators, Philadelphia, PA, 11-12 October 1989.

MEDIA ITEMS: (Group D.)

D1. Associate Press (national), May 1985, various titles including "New Exam Takes Guessing Out of Multiple-Choice Test."

Created awareness nationwide, including commentaries by radio announcers.

D2. Pantagraph (McLean County, IL), August 31, 1986, page A3, "ISU Professor Devises Testing Method to Make Guessing a Thing of the Past."

Profile of Dr. Anderson as professor, innovator and businessman.

D3. Business to Business (McLean County, IL) February 1987, pp. 24-27, "Computerization Takes Drudgery Away From Teachers."

Profile of Dr. Anderson's activities.

D4. Technological Horizons in Education (T.H.E.) Journal (national), March 1987, pp. 46-47, "Teacher's Testing Method Eliminates Guessing on Machine Scored Exams."

This is perhaps the best single-page discussion of MDT methods and usage prior to 1987.

BIBMDT.LAS

Brief Introduction to Computer-Assisted Reconsiderative Scoring: A New Method for Partial Credit and Multiple Correct Responses

Paul S. Anderson, Ph.D., Dept. of Geography-Geology, Illinois State University, Normal, IL 61761
MDT Corporation, 107 Broadway, Normal, IL 61761 (309) 452-6388

Reconsiderative test scoring occurs when a student's response to a multiple choice or completion (fill-in-the-blank) question is read by the teacher prior to the determination of correctness and point value. Historically, reconsiderative methods have been used extensively in manual scoring of completion-style questions. Examples include full or partial credit for synonyms, misspelled responses, numeric calculations with minor errors, and incomplete answers. The teacher has complete control. The method incorporates the best aspects of objective and subjective assessments because the teacher uses his/her content knowledge and professional judgement when making decisions after reading the student's response.

Computer-assisted reconsiderative scoring uses a special item analysis tabulation of responses (Figure 1). The actual word responses are shown on the computer screen. By moving the computer cursor (place designator) down the lines in the column called "Points", the teacher can enter from the keyboard the number of points each response should receive. Then the microcomputer program called RECON (tm) does all the re-scoring to allocate points to each student.

The ability to generate tabulations showing the actual word responses to completion questions has been available since 1983. (See The MDT Innovation, 1987, by Anderson on ERIC microfiche ED 307 287.) An alphabetized "answer bank" with up to 1000 terms is a distinguishing component. Each term has an MDT multi-digit identifier that can be marked on answer forms for machine reading. The "answer bank" lists can be prepared by individual teachers or by progressive textbook publishers and professional societies.

Three independent research studies in 1988 (abstracts available) support the use of this relatively new MDT testing method. The research results indicate 1) strengthened face validity, 2) appropriate use with problem solving and diagnoses, 3) improved identification of marginal examinees, 4) acceptance by students, and 5) evaluative power equal to that of fill-in-the-blank questions. Other operational advantages include ease of writing questions, reduction of bias, and improved feedback to students and teachers. The MDT method offers greater academic rigor because it eliminates recognition associated with multiple choice cued responses. The MDT multi-digit method is essentially machine-scored fill-in-the-blank assessment, especially when used with the RECON capabilities that became available in 1990.

In addition to the major advantages of reconsiderative scoring and awarding partial credit, RECON also permits the following innovative features: questions requiring multiple responses (Figure 2); calculated numeric responses (Figure 3); "multi-letter" responses in which every combination (such as ACD) of the five cued foils may be a response for full, partial or no credit (Figure 4). Future enhancements can accommodate decimal points, multiple steps, graphical responses and much more.

All indications are that computer-assisted reconsiderative scoring can be easily understood and utilized by teachers of all subject matter in schools ranging from upper elementary through college, medical schools and vocational training. By allowing teachers to have subjective input where appropriate, reconsiderative scoring should stimulate greater usage of higher order questions. When progressive teachers utilize these new capabilities and when "education corporations" (for scanners, textbooks and tests) provide the essential sponsorship for research and usage, the MDT and RECON innovations could make a major contribution to the improvement of American education.

Code	Answer	Pts.	Sub.	Freq.	Percent
000	(Blank)	0	0	2	2.4
186	Canada	2	7	1	1.2
307	England	1	7	6	7.1
325	France	0	0	1	1.2
328	Great Britain	0	0	4	4.7
412	Ireland	0	0	3	3.5
537	United Kingdom	2	7	67	78.8
562	Yugoslavia	0	0	1	1.2
	TOTALS			85	100.0

Figure 1: Example of a microcomputer display for reconsiderative scoring. A class of eighty-five students could be asked a free-response question: "Elizabeth II is the queen of what country?", with the following on-screen item analysis. The point values are designated by the teacher when moving the cursor up and down in the "Points" column.

Code	Response	Pts.	Sub.	Freq.	Percent
004		0	0	2	6.5
031		0	0	1	31.0
042		1	4	4	12.9
043		2	4	13	41.9
044		2	4	7	22.6
048		1	4	2	6.5
114		0	0	1	3.2
430		0	0	1	3.2
	TOTALS			31	100.0

Figure 3: Reconsiderative scoring of numeric responses. A science laboratory exercise about measurement could ask the following question: "To the nearest whole gram, what is the weight of the yellow precipitate in experiment 3?"

Code	Response	Pts.	Sub.	Freq.	Percent
017	Actinomycosis	0	0	18	22.5
102	Bacterial Meningitis	3	9	56	70.0
103	Bacterial Meningo-encephalomyelitis	0	0	21	26.3
806	Rabies	3	9	77	96.3
907	Tetanus	0	0	5	6.3
916	Thromboembolic Meningo-encephalitis	3	9	62	77.5
	TOTALS			560	700.0

Figure 2: Complex medical diagnosis question: Questions 2-8: Give seven differential diagnoses for the following case. Data: Hereford, 650 lbs., feedlot steer, vaccinated (IBR/BVD/PI3). Symptoms: Sudden onset of blindness, tremors, frothy salivation, opisthotonos, gets better, then gets worse. (Class size is 80 students, so 560 responses (7 x 80) are scored and tabulated.)

101 A	106 AB	116 ABC	126 ABCD
102 B	107 AC	117 ABD	127 ABCE
103 C	108 AD	118 ABE	128 ABDE
104 D	109 AE	119 ACD	129 ACDE
105 E	110 BC	120 ACE	130 BCDE
.....	111 BD	121 ADE	.....
	112 BE	122 BCD	131 ABCDE (all)
	113 CD	123 BCE	132 (none of them)
	114 CE	124 BDE	
	115 DE	125 CDE	
	.....	.....	.....

Figure 4: Thirty-two possible combinations of five letters, each with an MDT multi-digit number. Any question with up to five alternatives labeled A, B, C, D, and E could be used with this special MDT list for "multi-letter" responses. For example: "Which of the following characteristics is/are commonly associated with [whatever topic or situation the teacher chooses to present]: A) ...[word, phrase, sentence or even paragraph]... B) ... C) ... D) ... E) ...